Application of Process Mining in Managerial Accounting: A Case Study of an Investments and Securities Firm

Jason Jihoon Ree

Technology Management Major
School of Technology Management
Graduate School of UNIST

2011

Application of Process Mining in Managerial Accounting: A Case Study of an Investments and Securities Firm

Jason Jihoon Ree

Technology Management Major
School of Technology Management
Graduate School of UNIST

2011

Application of Process Mining in Managerial Accounting: A Case Study of an Investments and Securities Firm

Jason Jihoon Ree

Technology Management Major
School of Technology Management
Graduate School of UNIST

Application of Process Mining in Managerial Accounting: A Case Study of an Investments and Securities Firm

A thesis
submitted to the School of Technology Management
and the Graduate School of UNIST
in partial fulfillment of the
requirements for the degree of
Master of Science

Jason Jihoon Ree

07. 18. 2011
Approved by

Major Advisor
Minseok Song

Application of Process Mining in Managerial Accounting: A Case Study of an Investments and Securities Firm

Jason Jihoon Ree

This certifies that the thesis of Jason Jihoon Ree is approved.

07. 18. 2011

Thesis supervisor: Minseok Song

Kwanho Kim

Han-Gyun Woo

Abstract

Business process mining techniques use event logs recorded from information systems to extract and discover useful process and organizational information. Over the past couple of decades, many new and powerful process mining techniques have been developed by researchers and software vendors. Although numerous case studies demonstrating the applicability of process mining have been emerging in literature, there has yet to be an application of process mining in the financial sector. In this paper, we explore the applicability of process mining of an investments and securities firm, more specifically investigating its managerial accounting processes. Using an assortment of process mining techniques available in the ProM framework, we examine and discover clear differences between the AS-IS model stated by the financial firm and the process model extracted from the event logs. This research aims to accomplish the following: a) add to the current process mining application literature by examining the applicability of process mining in the financial sector, b) utilize various process mining techniques to observe and assess process information in the specific managerial accounting case study of the monthly profit and loss computation process, c) identify the strengths of process mining and how they can supplement the weaknesses of business process reengineering, and d) address the possibility that event log data with insufficient case sizes needs to be addressed differently than past approaches with data with sufficient case sizes.

Contents

List of Figures

List of Tables

I. Introduction

From total quality management to organizational transformation and from right-sizing to continuous improvement, there exists an extensive list of approaches with the common goal to change how *business processes* are conducted in order to improve organizational performance. Among these approaches, business process reengineering (BPR) emerged as the "hottest management concept since the quality movement" (Byrne, 1993) in the early 1990s. Business process reengineering is defined as the fundamental rethink and radical redesign of business processes to generate dramatic improvements in critical performance measures, such as cost, quality, service and speed (Hammer & Champy, 1993). Furthermore, Hammer and Champy (1993) stated that BPR should not be restricted by the existing situation. In other words, BPR promises to transform organizations by fundamentally altering their core processes, thereby achieving radical improvements in performance. Despite the stories of dramatic success in business process improvements via BPR, such success stories stand in contrast to the many studies reporting supposed high failure rates associated with BPR (Larsen & Myers, 1997).

Motivated by the many unsuccessful outcomes of BPR, there have been numerous studies, which examine the factors and possible causes of BPR failure. Our research aims to take a different standpoint from the majority of past BPR failure research by focusing on how the assessment methods of information and data regarding the business process may be lacking, and thereby possibly leading to unsuccessful BPR implementations. One common procedure of BPR involves assessing and gathering current information and knowledge regarding the status quo of the organization and its processes through employee interviews and surveys. Sourced from such approaches, BPR can be subject to misleading information, biases, and so forth. Furthermore, there can be certain types of information unobtainable through these approaches. Such data and information assessment methods of BPR may be one of the many reasons for causing such low success rates.

Hence, our belief is that process (workflow) mining can be a significant *supplementary* approach to existing methods by assessing and discovering the current status of organizational processes (AS-IS), detect and measure unwanted deviations, and to provide new perspectives, comparisons, and extensions of processes with performance assessments (TO-BE). We have confidence that process mining can help overcome the shortcomings in the data and information assessment methods (e.g. acquiring AS-IS system descriptions, identification of improvement opportunities, etc.) in implementing BPR by providing more concrete event log data-supported measures and insights. Process mining aims at the extraction of organizational knowledge regarding processes from event logs attained from numerous varieties of systems, such as information systems and embedded systems (van der Aalst, 2003). From event logs, the information of the start and completion of process steps and procedures including related data (e.g. originators and resources) can be potentially acquired

(Mans et. al., 2008). The contributions in the process mining domain since the mid-1990s have provided newly created algorithms and useful process analysis tools; and following each new process mining technique, applications of these techniques in various fields were researched and analyzed. In short, the domain of process mining research can be categorized into its (1) techniques and (2) applications.

This study focuses on the applicability of process mining in the financial sector. With the successes of process mining applications in the service industry (van der Aalst, 2007), healthcare sector (Maruster et. al., 2002; Mans et. al., 2008), and so forth, this study aims to demonstrate the application of process mining in a new field, the financial sector. We are motivated to explore the applicability of process mining of new realms, in part, considering van der Aalst et. al. (2007)'s statement that with the advancements in process mining techniques, applying process mining in a wide variety of practical situations is high in priority; along with the lack of an application of process mining in the financial domain. Moreover it is noteworthy that the vast majority of prior process mining application research has only been applied to rather structured processes (Rozinat et. al., 2009). This study aspires to add to the growing process mining application literature with an application in the financial domain, which tends to contain unstructured, non-linear, complex processes.

There exist innumerable processes that transpire in the financial sector. However as an initiation in examining the financial domain using process mining, we concentrate on the managerial accounting aspect of an organization in this study. In any business domain, managerial accounting of an organization is an essential and determining element in an organization's success and prosperity. By definition, managerial accounting is the provisions and use of accounting information to managers within organizations, to provide them with the basis to make informed business decisions that will allow them to be better equipped in their management and control functions (Garrison & Noreen, 2009). It is a widely accepted notion that managerial accounting potentially plays a crucial role in bringing success to firms of various sizes and business domains by assessing the firm's current status and further determining future directions for the firm. We examine a case study involving managerial accounting processes of an investment securities firm. The firm desires to improve its important processes by undergoing process reengineering, however they are skeptical of the costs and success of BPR. For our research, we analyze the monthly profit and cost computation process of the firm. We believe the use of process mining can be utilized as a significant tool in analyzing managerial accounting processes due to the fact that managerial accounting processes are not as structured as it might be portrayed. The data was obtained from the firm's database of the managerial accounting process records. Furthermore, the AS-IS process model was acquired through an interview with experts of the firm.

By using process mining techniques, the goal of this study is to discover the real process model of the process and pinpoint current flaws and shortcomings in the process by comparing the AS-IS process with the process model extracted through process mining of the event logs. However quite interestingly, the investments and securities firm provided an event log with only three cases. Most research thus far has only applied process mining where the processes contain an abundant number of cases. Thus this study addresses the possible need for a different approach needed to tackle applications of process mining in the context of less than sufficient case sizes. We believe such research will be a significant contribution to the application literature of process mining because the proportion of data with insufficient cases is most likely to become more common as the applicability of process mining is further researched. Furthermore, the methods in which to study the applicability of process mining under such circumstances must be approached with caution due to, for example, the data being more prone to noise or the extraction of incorrect process models via process mining techniques.

In summary, this research aims to accomplish the following: a) add to the current process mining application literature by examining the applicability of process mining in the financial sector, b) utilize various process mining techniques to observe and assess process information in the specific managerial accounting case study of the monthly profit and loss computation process, c) identify the strengths of process mining and how they can supplement the weaknesses of business process reengineering, and d) address the possibility that event log data with insufficient case sizes needs to be addressed differently than past approaches with data with sufficient case sizes.

II. Related Work

2.1 Business Process Reengineering

To be a successful organization, companies need to work as a team and all the functional areas of the business need to be properly integrated, with each understanding the importance of cross functional processes. As the basis of competition changes from cost and quality to flexibility and responsiveness, the value of process management has been recognized. The role that process management can play in creating sustainable competitive advantage was termed Business Process Reengineering, and was first introduced by Hammer (1990) and Davenport and Short (1990). By definition, BPR is the fundamental rethinking and radical redesign of business processes to achieve dramatic improvements in critical contemporary measures of performance such as cost, quality, service, and speed (Hammer & Champy, 1993). Additionally, authors have already pinpointed that many different approaches exist—for example, Hess and Oesterle (1996) evaluate 12 methods–which makes it difficult to define what exactly BPR constitutes (Braganza andMyers, 1996; Choi and Chan, 1997; O'Neill and Sohal, 1999) and to define its methodology (Childe et al., 1994). Al-Mashari and Zairi (2000) conclude that all the definitions emphasize redesigning business processes using a radical IT-enabled approach to organizational change. BPR perceives business processes as horizontal flows of activities, while most organizations are formed into vertical functional groupings (Dekkers, 2008). Dekkers (2008) also stated that BPR works in two ways: it enables an organization to cope with external changes and it enhances competitiveness affecting the competitive landscape.

Since its introduction in the 1990s, BPR has been one of the more popular methodologies used to redesign existing processes. But despite its popularity, BPR has been subject to criticisms and doubts (Clemons, 1995). Various studies have examined the success of BPR implementation in corporations and organizations, where Hammer and Champy (1993) estimated that only 30-50% of the efforts succeed, while Jarrar and Aspinwall (1999) found only 25% of reengineering cases reach acceptable performance improvements. More interestingly, Sabherwal et al. (2001) found that alignment between strategy and structure is hardly achieved, and that a redesign is often inhibited by cultural and structural inertia. Most of all, many writers attribute the failure of reengineering efforts to leadership, culture, change management, etc. (e.g. Braganza and Myers, 1996; Campbell and Kleiner, 1997; Choi and Chan, 1997; Drago and Geisler, 1997). Some like Bryant and Chan (1998) attribute the low success rate to BPR as a covert way for downsizing. On the contrary, we believe that the there may be faults in the data and information assessment methods (e.g. acquiring AS-IS system descriptions, identification of improvement opportunities, etc.) in implementing BPR; and suggest process mining as a auxiliary tool in the assessment and implementation of BPR by providing more concrete event log data-supported measures.

2.2 Managerial Accounting

Managerial accounting provides the essential data with which the organizations are actually run. Managerial accountants prepare a variety of reports. Some reports focus on how well managers or business units have performed-comparing actual results to plans and to benchmarks; some reports provide timely, frequent updates on key indicators such as orders received, order backlog, capacity utilization, and sales; while other analytical reports are prepared as needed to investigate specific problems such as a decline in the profitability of a product line (Garrison & Noreen, 2009).

2.3 Process Mining

Process mining has proven to be a valuable approach that provides new and objective insights into the way business processes are actually conducted within organizations (Rozinat et. al., 2008). Since the mid-1990s, process mining and its techniques has been thriving through the contributions of the researchers in the field.

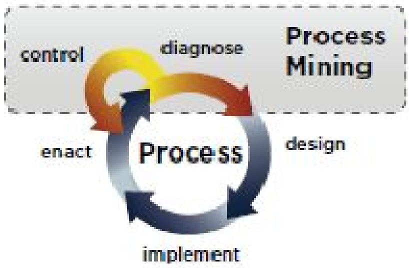


Figure 1. Position of process mining in Business Process Management (Rozinat, 2010)

In general, process mining can be categorized in the broader field of Business Process Management (BPM) (Rozinat, 2010). As seen in Figure 1 (the life cycle of BPM), process mining falls into the last phase of BPM, *process diagnosis*, where the running process is analyzed to identify problems, or to find ideas for improvement (Rozinat, 2010); thus enabling both direct process control and *process redesign*.

The idea of process mining is to discover, monitor, and improve real processes (i.e., not AS-IS processes) by extracting knowledge from event logs (Mans et. al, 2008) which may originate from all kinds of systems (i.e. enterprise information systems) (Boumen et. al., 2006) to use it for a detailed analysis of reality (Song & van der Aalst, 2008). One condition is that the system must produce event logs recording the activities or actual behavior of the processes. In general such event log generating systems are classified under Process-Aware Information Systems (PAISs) (Dumas et. al., 2005).

Thus far, process mining research has expanded into two general areas. One important area being the creation of new algorithms and process mining techniques, such as Alpha miner (van der Aalst et.

al., 2004), Heuristic miner (Weijters and van der Aalst, 2003), Alpha++ miner (Wen et. al., 2006), Genetics miner (Alves de Medeiros, 2006), Trace Clustering (Song et. al., 2008), Fuzzy Clustering (van Dongen & Adriansyah, 2009), Social Network mining (van der Aalst and Song, 2004), and many more.

The second area can be generalized as explorations and utilizations of process mining in the contexts of case studies, such as software development processes (Urena Hinojosa, 2008; Rubin et. al., 2007), hospital and healthcare processes (Maruster et. al., 2002; Mans et. al., 2008), manufacturing (Rozinat et. al., 2009), case handling systems (Athena, 2002), supply chains (Maruster et. al., 2002), and web services (Arkin et. al., 2005; Kavantzas et. al., 2004). Also there has been more recent research applied to auditing (van der Aalst et. al., 2010; 2011) and product usage monitoring (Funk et. al., 2009).

Figure 2 shows the general conceptualization of process mining from event logs recorded from information systems and the three classes of process mining. There exist three fundamental classes of process mining techniques: 1) *discovery*, 2) *conformance*, and 3) *extension.*

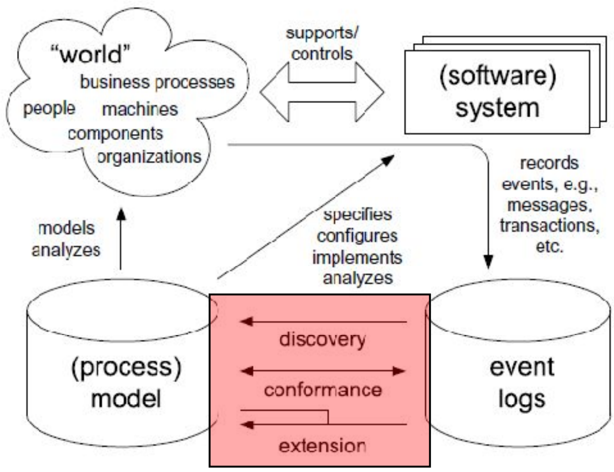


Figure 2. Three classes of process mining: discovery, conformance, and extension

Discovery

By tradition, process mining research and literature has been revolving around discovery, where deriving information about the original process model, the organizational context, and execution properties from enactment logs (Mans et. al, 2008). In the discovery phase of process mining, process mining techniques, such as α-algorithm mining and heuristic algorithm mining, are utilized to construct visual representations (i.e., Petri nets) of real processes extracted from event logs. In other words, process discovery algorithms automatically construct process models based on the behavior that was observed in the event log (Rozinat et. al., 2010). However, discovery is not limited to only the extraction of control flow and process models; recent process mining techniques are becoming more focused on other perspectives, for example the organizational perspective, performance perspective, and data perspective (Mans et. al, 2008). This discovery phase plays an important role in setting the foundation for the analyses of real processes.

Conformance

In the conformance phase of process mining, the process models extracted from reality (i.e., event logs) are compared with an AS-IS model. Conformance checking may be used to detect deviations, to locate and explain these deviations, and to measure the severity of these deviations (van der Aalst & Günther, 2007). Conformance checking techniques evaluate the relation between process models and reality presented in from of event logs through orthogonal dimensions of conformance (i.e., fitness, precision, generalization, and structure) (Adriansyah et. al., 2010).

Extension

In the extension phase of process mining, the process model is extended with new aspects or perspectives to enrich the model with the data in the event log (Mans et. al., 2008). Some examples of extension are illustrating the bottlenecks in a process model by analyzing the event log or utilizing decision point analysis to detect data dependencies that affect the routing of a case (i.e., how data affects the choices made in the process based on past process executions) (Song & van der Aalst, 2008).

III. Research framework

For our analysis we follow a process mining framework as illustrated in Figure 3 constructed by Song et. al. (2010). The framework starts by 1) scoping and acquisition of the data, then 2) preparation of the data. Next 3) a quick scan and preprocessing is performed, followed by 4) conducting process mining techniques of the data. Afterwards 5) further analysis is performed. Finally f) conclusions are made based on the analyses.

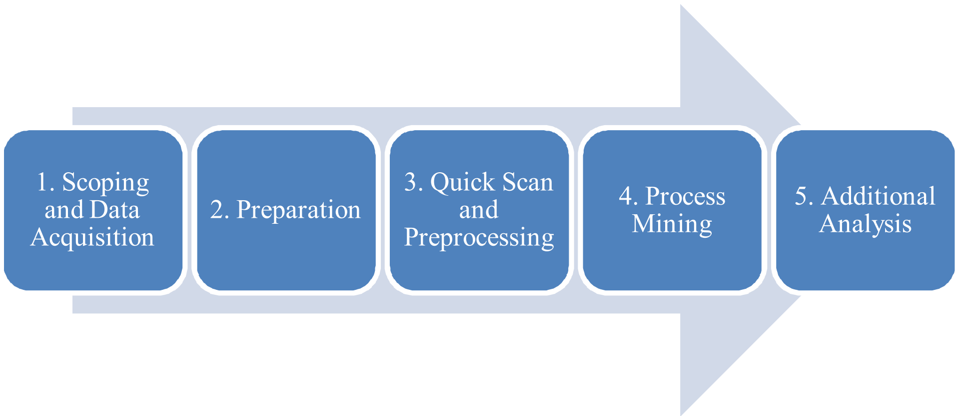


Figure 3. Process Mining Framework

The following list describes the details of each stage of the process mining framework (Song et. al., 2010):

1. Scoping and Data Acquisition

- Selection of target process and its relevant information system
- Selection of the quantity of data
- Aggregation data from relevant information system

2. Preparation

- Conversion of data into *Mining* XML (MXML) format for analysis preparation

3. Quick Scan and Preprocessing

- Collection of preliminary information for obtaining analysis direction and guidelines
- Deduction of unstable cases which may create unnecessary noise in process analysis
- Selection of "sound" cases for thorough analyses

4. Process Mining

- Utilization of process mining techniques for analyses in the organizational perspective, performance perspective, and workflow perspective

5. Additional Analysis

- Utilization of filtering and clustering techniques for further analyses

IV. Case Analysis

The applicability of process mining in the financial sector will be demonstrated through the analyses of the managerial accounting data. More specifically, the monthly profit and loss computation process of the firm will be explored. This study will focuses mainly on the discovery and conformance aspects of process mining with the addition of social network analysis, performance analysis, and so forth. After a short-description of the log data recorded in the database, the analyses using process mining techniques of the event log data are explicated in this section.

4.1 Log Description

The event log data collected was from the information system database of the assessment of monthly profit and loss process from April of 2010 to June of 2010. The log of the investments and securities firm contains a large quantity of distinct activities performed by individuals in different departments of the company. Preprocessing of the data was performed to consider events at the department level. Through a preliminary examination of the data after preprocessing the event log, we found that the data consists of three cases, fourteen activities, and 366 events shown in Table 1.

Table 1. Log description

Categories	Values
Number of Cases	3 (April, May, June 2010)
Number of Activities	14
Number of Events	366
First timestamp	Apr 16, 2010 10:13:39 AM
Last timestamp	Jul 26, 2010 4:07:23 PM

Before analyzing the event log, observing the AS-IS model, shown in Figure 4, obtained through interviews with the experts is important in gaining general information regarding the process. There exist seven departments: (i) the Business Management Department, (ii) Human Resources Department, (iii) Payment Services Department, (iv) General Affairs Department, (v) Training Department, (vi) E-Business Department, and (vii) Sales Department.

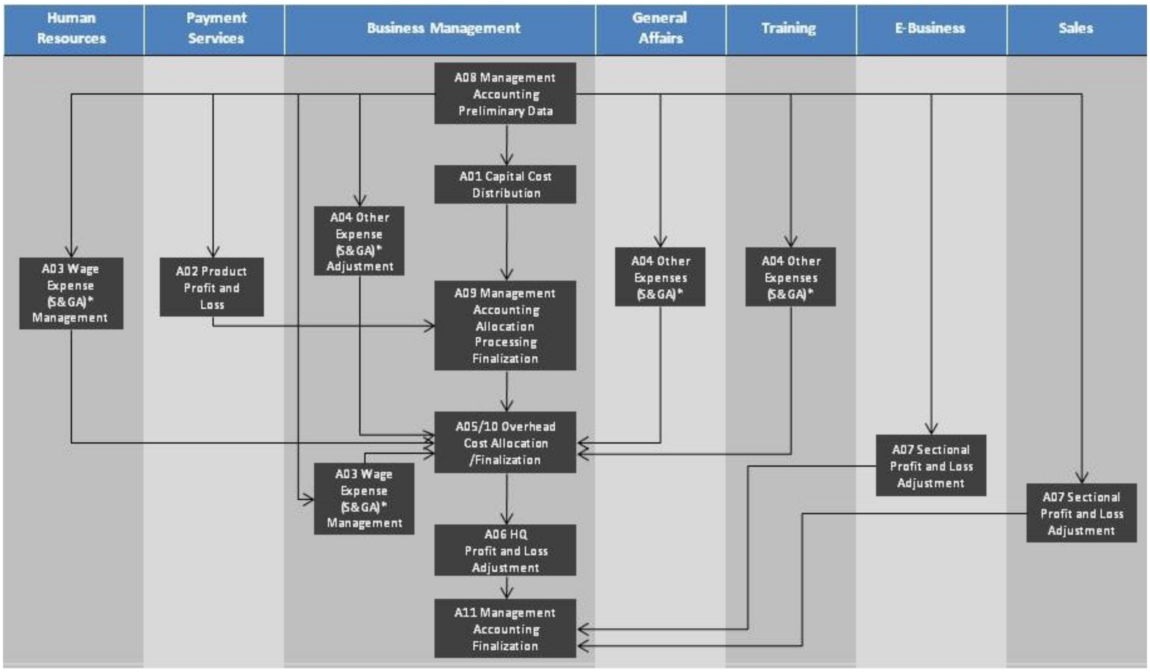


Figure 4. AS-IS model of the process

The majority of activities are conducted by the business management department. According to the AS-IS model, the business management handles the core essential tasks, and is intertwined with all other departments; the other departments conduct more minor tasks, and interact and provide relevant information to aid the business management department in assessing the overall computation (i.e. expenses, profits, wages, etc) of the firm. Less vital tasks such as wage expense, sectional and product profits and loss computations as well as other selling and general administrative expense computations are handled by these supportive departments in this process. From a managerial accounting perspective, such departments play a supportive role to the business management department, where the core tasks are performed. A more specific analysis of the handover of work will be discussed later by means of social network analysis. From Figure 4 alone, it can be observed that the monthly profit and loss computation process is a simple and straight-forward process. Using process mining techniques, we examine whether the AS-IS model is an accurate representation of the actual behavior of the process recorded in the event log.

4.2 Mining

This section presents the analyses of the case study of the application of process mining to the monthly profit and cost computation process of an investment securities firm. The process of monthly profit and cost computation of the investment securities firm contains three cases, in which each case represents the processes recorded each month between April, 2010 and June, 2010.

4.2.1 Control Flow Perspective

For the purposes of process discovery, the AS-IS model was also coded into Petri net format for analysis using the ProM framework in Figure 5. Figure 6 shows a considerably simple, straight forward AS-IS model, as the experts of the firm envisioned their monthly profit and loss computation process to represent.

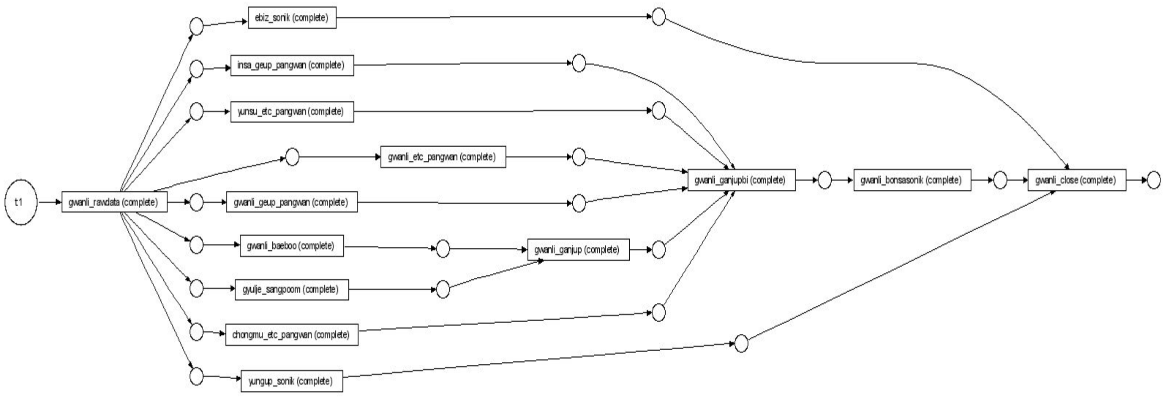


Figure 5. Petri net model of AS-IS process

The analysis of the event log begins with control flow mining, a powerful and constructive technique, which automatically derives process models from process logs. The process model that is generated presents a model of the actual process as observed through real process executions (Mans et. al., 2008). This model provides insight into actual work flow of activities in the process for analysts and managers. There exist several process mining algorithms such as α-mining algorithm (van der Aalst et. al., 2004) and heuristic mining algorithm (Weijters and van der Aalst, 2003).

In this study, we utilize both algorithms, each technique for different purposes. α-mining algorithm bestows a model in Petri net format, which then can be utilized to perform further analyses techniques. For example, certain process mining techniques like conformance checking and performance analysis require Petri net models for analysis. Heuristic mining algorithm is a stronger candidate in obtaining a model that is robust against noise and exceptions, and enables users to focus on the main process flow instead of every detail of the behavior present in the process log (Weijters et. al., 2003).

Figure 6 shows the process model for all cases using α-mining algorithm. As seen in Figure 6, the α-mining algorithm model is somewhat spaghetti-like, thus too complex to interpret easily. There are several methods in which to simplify and avoiding the problem of model complexity, such as clustering or dissecting the log to pinpoint specific areas of interest within the model. Such methods were found successful in the analyses of complex, unstructured models in the application of process mining in the healthcare domain (Maruster et. al., 2002; Mans et. al., 2008). However, in comparison, the process model extracted from the event log in this case is not as complex, nor were there enough cases for clustering techniques to be found effective. Furthermore simplification methods such as clustering can be subject to losing information regarding the important tasks and interrelated connections within the process. This was found to be factual when clustering techniques were performed on our data for testing purposes. Thus it was decided that filtering techniques (start-task filter; Figure 6(ii)), built in the ProM framework, was a more plausible solution in the early stages of our analysis.

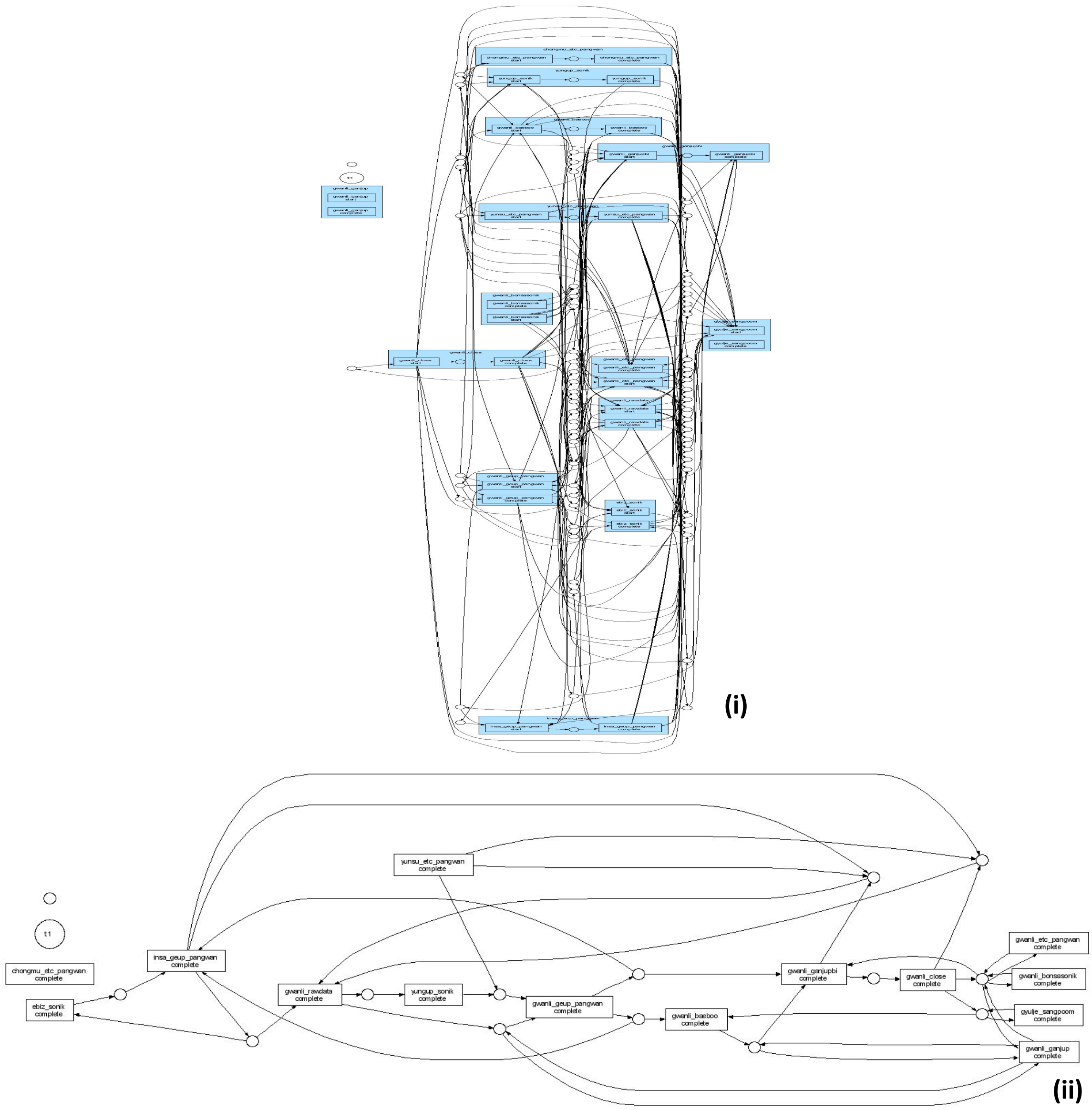


Figure 6. α-mining algorithm model

(i) α-mining algorithm model without filter; (ii) α-mining algorithm model with start-task filter

Figure 7 shows the heuristic mining algorithm model. Although the model extracted from using the heuristic mining algorithm in Figure 7(i) is much simpler than the α-mining algorithm model in Figure 6, filters were also utilized for this model resulting in a simplified model (start-task filtered out; Figure 7(ii)).

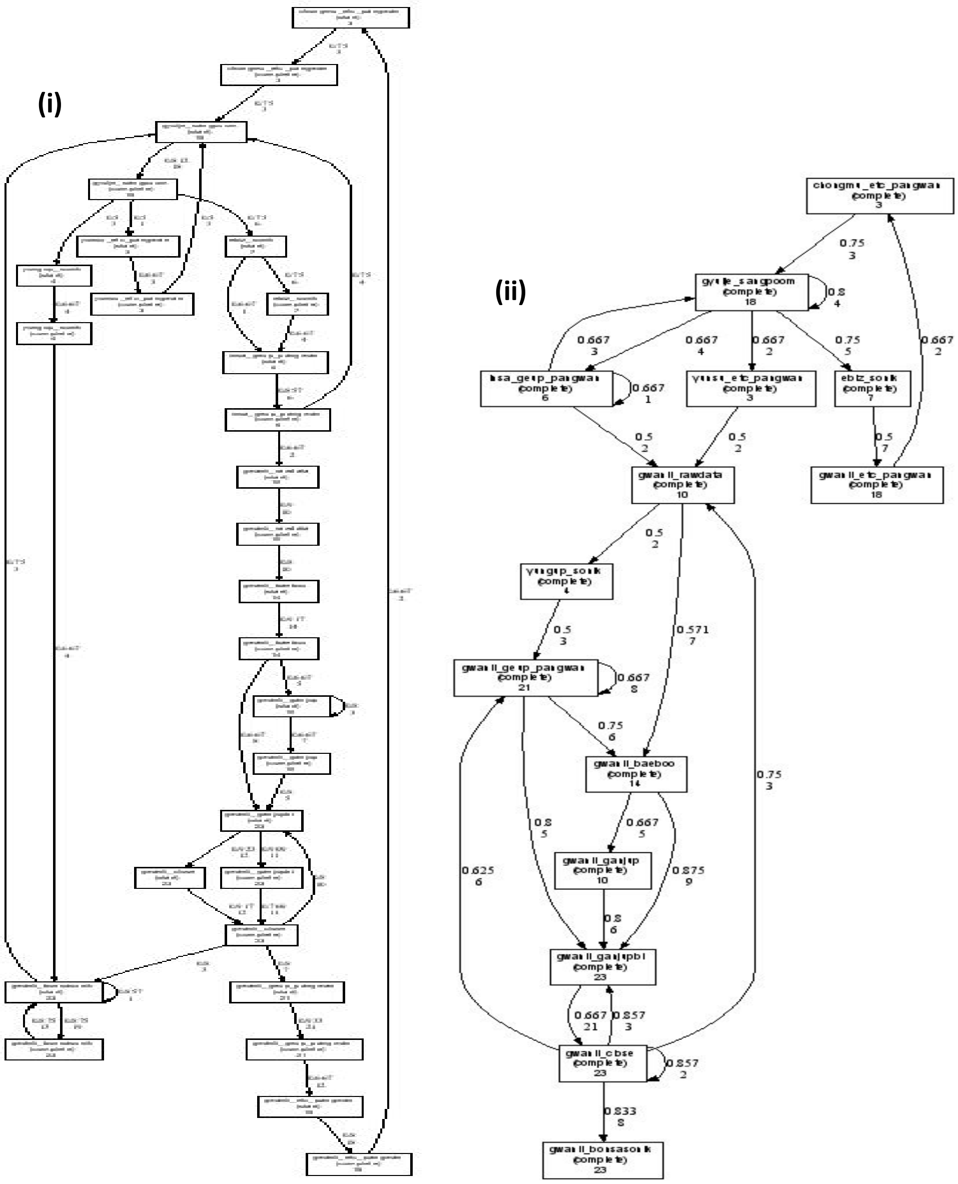


Figure 7. Heuristic mining algorithm model
(i) Heuristic mining algorithm model without filter; (ii) Heuristic *mining algorithm model with start-task filter*

In reference to Figure 6, the presence of isolated tasks and places can be seen in the α-mining algorithm model. Comparatively, the heuristic mining algorithm model is obviously more robust and provides more clear and significant information regarding the process visually.

To obtain a better visual representation of the differences between the AS-IS model and the model extracted from heuristic mining algorithm, Figure 8 shows the overlay of the workflow extracted from the event log onto the AS-IS model.

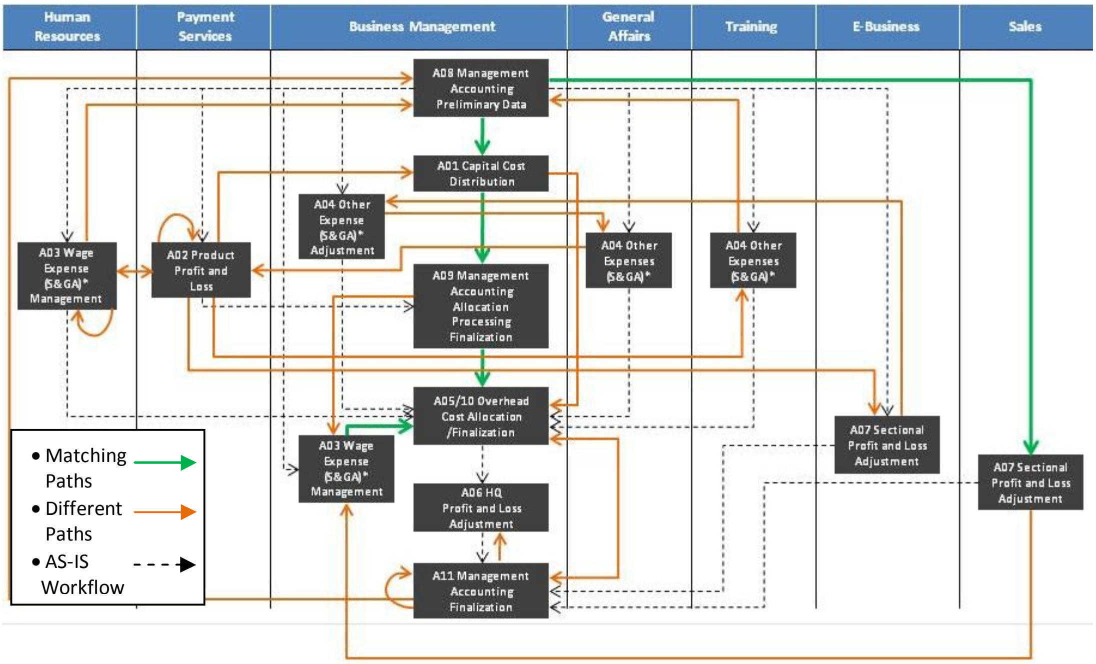


Figure 8. Overlay of actual workflow on the AS-IS model frame

There only exist five paths (green arrows) from the AS-IS model that was found to match the actual workflow. The dotted arrowed lines represent the original AS-IS workflow paths which were not found to match the workflow paths of the event log. Moreover twenty paths (orange arrows) were also drawn to show that the process is much more complex and dissimilar than the company experts had pictured. Additionally, the extracted process was isolated from Figure 8 for further analysis of the actual workflow of the monthly profit and loss computation process in Figure 8.1. This figure is significant in a sense that the AS-IS model illustrates a top-to-bottom workflow and by isolating the extracted model onto the frame, not only are the differences of workflow paths identifiable, any workflow paths that oppose the top-to-bottom 'rule' can be recognized as workflow paths that are apparent in actuality but not familiar to the organization and its managers. In other words, process mining can discover and differentiate such workflow paths that may give managers better insight of what processes actually transpire and exist within their organizations, which were previously difficult, if not impossible, to identify.

For example as shown in Figure 8.1, three workflow paths starting from events: a) *Human Resources*: 'Wage Expenses (Selling and General Administrative Expenses) Management', b) *Business Management*: 'Management Accounting Finalization', and c) *Training*: 'Other Expenses (Selling and General Administrative Expenses)', return back to *Business Management*: 'Management Accounting Preliminary Data' at the top of the process model.

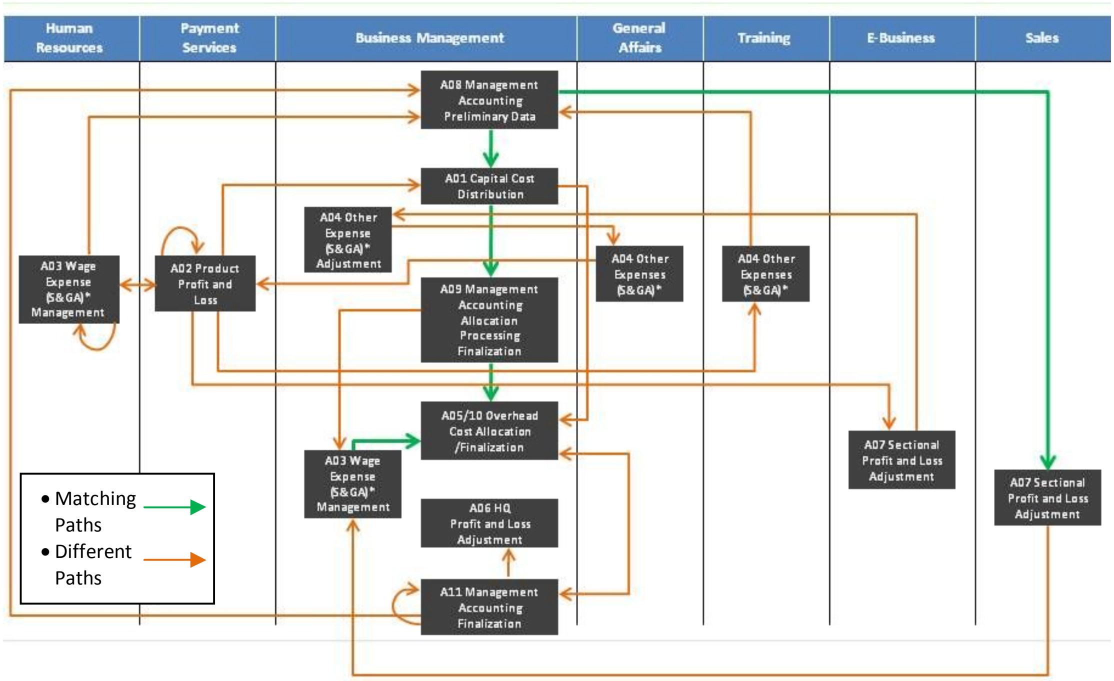


Figure 8.1 Actual workflow isolated on AS-IS frame

According to the AS-IS model, it is evident that such revisits to the 'Management Accounting Preliminary Data' activity is illogical and incongruous. Discovery of such workflow paths can bestow new information regarding the current process and, furthermore, whether or not such paths are abnormal, or whether they should be flagged as problems within the process for further examination.

4.2.2 Conformance Analysis

At this stage of the analysis, the question of which process model is most accurate comparatively was raised. If we assumed that the behavior observed in the log is what really happened and somehow representative for the operational process at hand, it is possible to compare the discovered model to the event log that was used as input for the discovery algorithm (Rozinat et. al., 2008). This essentially results in a conformance analysis problem (Rozinat et. al., 2008; Cook et. al., 1999). The most dominant requirement for conformance is the fitness metric (Rozinat & van der Aalst, 2007). Fitness can be defined as the extent to which the log traces can be associated with valid execution paths specified by the process model (Rozinat & van der Aalst, 2007). The possible fitness values (f) can range from 0.0 (corresponds to the case where the model and the log do not fit at all) to 1.0 (i.e., model and log fit to 100%) (Rozinat et. al., 2009). But unfortunately, a good fitness does not imply conformance. Thus another dimension used in conformance checking is appropriateness, which is the degree of accuracy in which the process model describes the observed behavior, combined with the degree of clarity in which it is represented (Rozinat & van der Aalst, 2007). However, in consideration of the fitness metric being a more dominant metric, for the purposes of this study, we will focus on the fitness metric alone when analyzing the models.

Conformance Analysis of AS-IS Model

Before analyzing the extracted models, we started analyzing the AS-IS process model first as a benchmark of what type measures are to be expected. Figure 9 illustrates the results of the conformance analysis for the AS-IS model matched with the event log. The fitness metric yielded 40.2%. This can be interpreted as only 40.2% of the event log data "fitting" the AS-IS model. Unfortunately this fitness value is quite low; leading us to question the possibilities of inaccurate information portrayed in the AS-IS model at the time it was acquired from the experts.

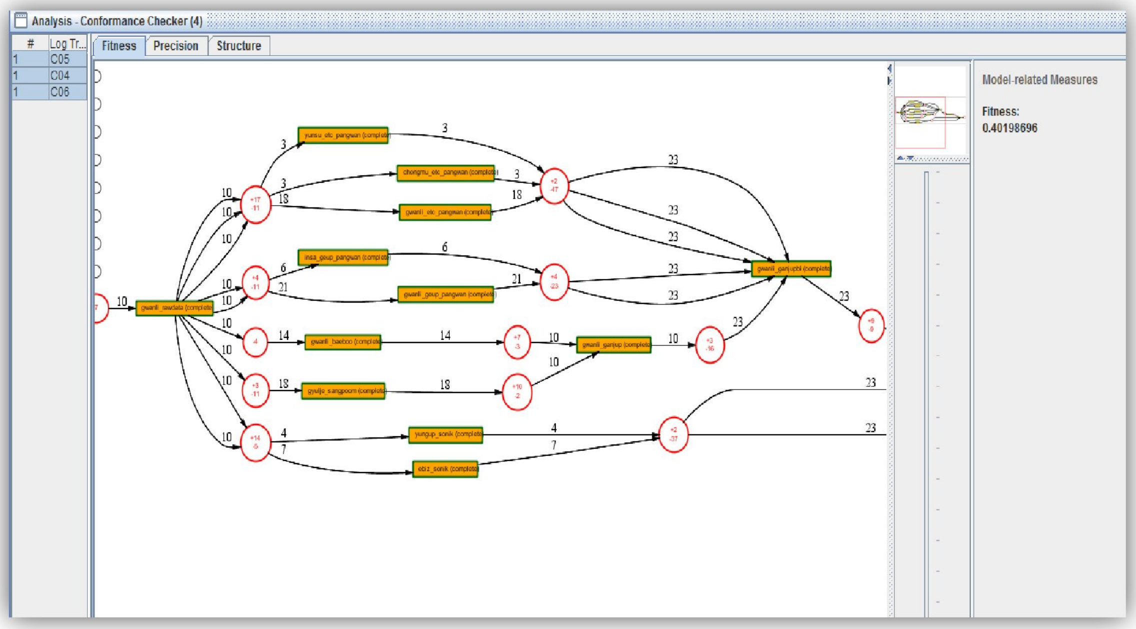


Figure 9. Conformance analysis of AS-IS model (Fig. 5)

After thorough examination of the event log data, we were able to find that, unlike the straight forwardness of the AS-IS model, a large majority of tasks were being repeated multiple times throughout the duration of each of the three cases. Some tasks were even found to be repeated one after another sequentially. Our interpretation of the differences between the AS-IS model and the behaviors shown in the event log was that information was omitted during the interview; in short, the experts were not aware of the severity of concealing the existences of task repetition. This exemplified one of the key problems in obtaining information via interviews (i.e., common assessment and diagnosis methods used in BPR) – the weakness of misleading and missing information.

As a result, in the AS-IS model, the tasks with repetitious behaviors from the observation of the event log were given loops, which allowed the AS-IS model to repeat tasks; and hopefully this would lead to a more robust and appropriate AS-IS model. Figure 10 shows the AS-IS model with the addition of loops for the repetitious tasks. The loops were coded into the original AS-IS PNML (Petri Net Markup Language) file as invisible tasks, or the grey squares in Figure 10. Invisible tasks are defined as the tasks that exist in a process model but not in its event log (Wen et. al., 2010). We utilize invisible tasks to modify models without the creation of new tasks, which can influence the authenticity the event log.

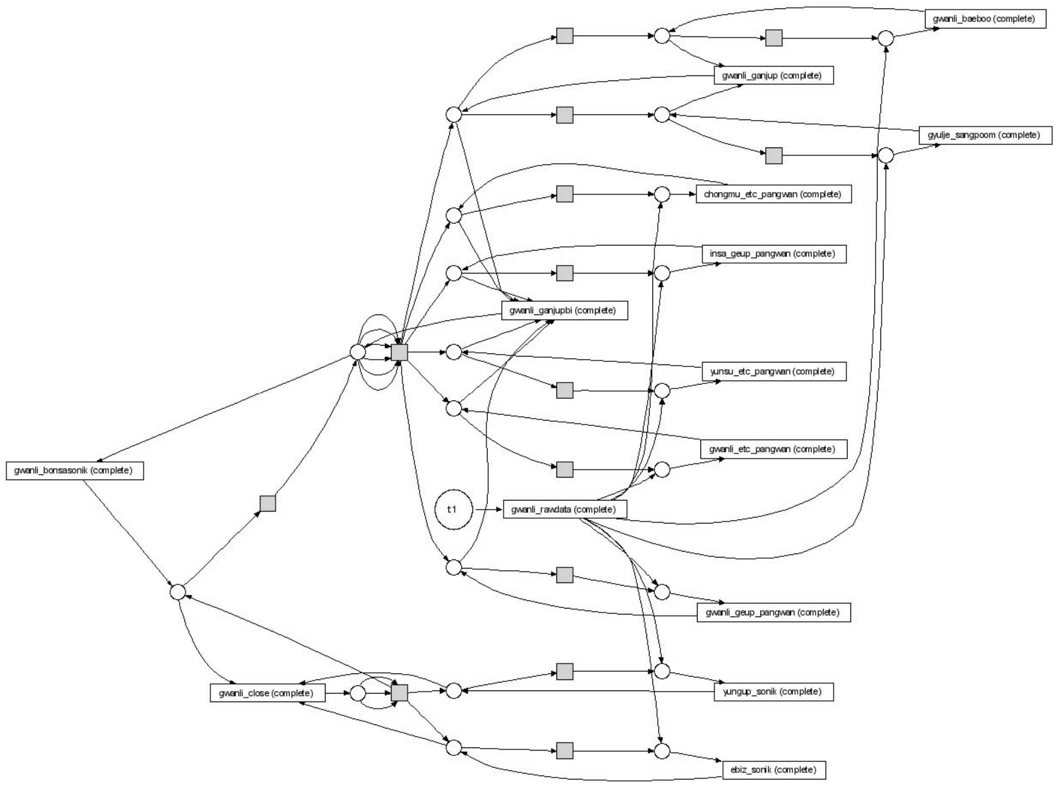


Figure 10. AS-IS model with loops

After modifying the AS-IS model, Figure 11 shows the conformance analysis, yielding a fitness of 60.7%. This is a vast improvement in fitness than the original AS-IS model.

In summary, this improvement in fitness of the AS-IS model can be explained by the lack of information provided by the experts during the interview procedure to construct the AS-IS model, thus the original AS-IS model was a misrepresentation of the assumed process conceptualized by the company experts. Although the fitness metric did show a promising increase, it is still considerably low. This occurrence of low fitness might be due to the fact that the task flows are revisiting back to the tasks in the earlier stages from the latter part of the model; and the AS-IS model does not have any means that structurally allow for such behaviors to occur. Such incidences imply that, not only is the monthly profit and loss computation process complex, but it is not as straight forward as the AS-IS model portrays it to be. We do not continue to modify the AS-IS model any further, since it would imply that we would be converging the AS-IS model to fit the event log with each additional modification. Although the AS-IS model describes the general and ideal flow of the process, we feel the addition of individual task loops was justified because of the fact that, in logs with low numbers of cases, extracted models may not interpret the repetition of certain tasks correctly; and by allowing the AS-IS model to include repetitious behaviors, the comparison between the AS-IS model and the extracted model would be more meaningful in a sense that the extracted models can be evaluated using the AS-IS model as a benchmark.

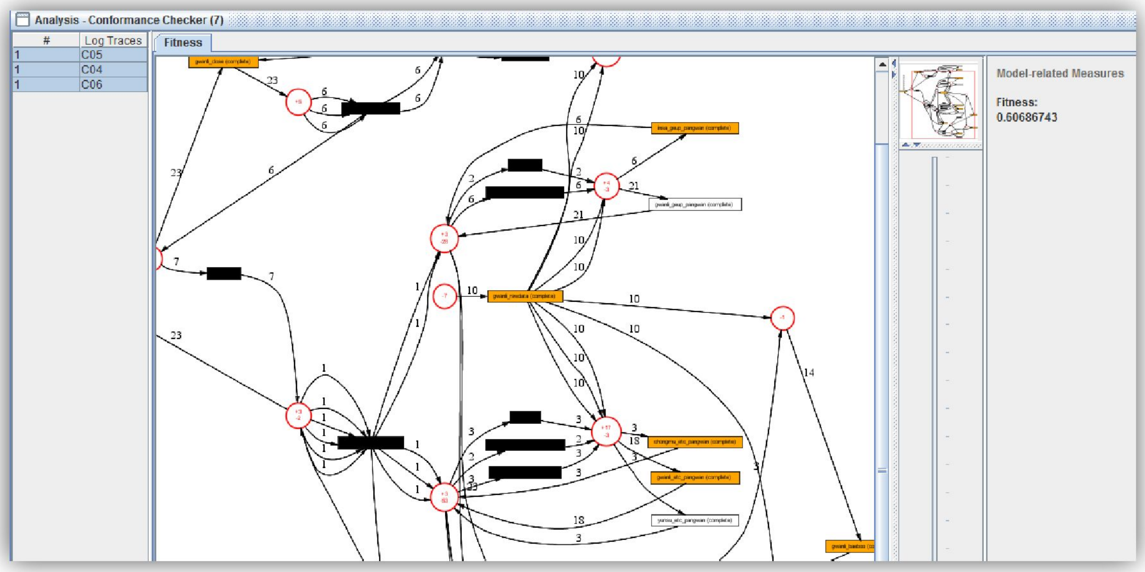


Figure 11. Conformance analysis of AS-IS model with loops (Fig. 10)

This procedure in discovering the best model for both the AS-IS model and the extracted models is important due to the fact that the extracted models yielded from process mining may lack validity because of the small number of cases present in this case study. Once the best models for the AS-IS and extracted models are acquire, proper comparative analysis can be conducted.

Conformance Analysis of Extracted Models

As an assessment of the fitness of the extracted models, conformance analysis was conducted. Figure 12 demonstrates the conformance checking plug-in applied to the model extracted from α-mining algorithm. The fitness of this model is 58.9%. Although the fitness metric is marginally higher than the AS-IS model fitness metric, this value is quite low also.

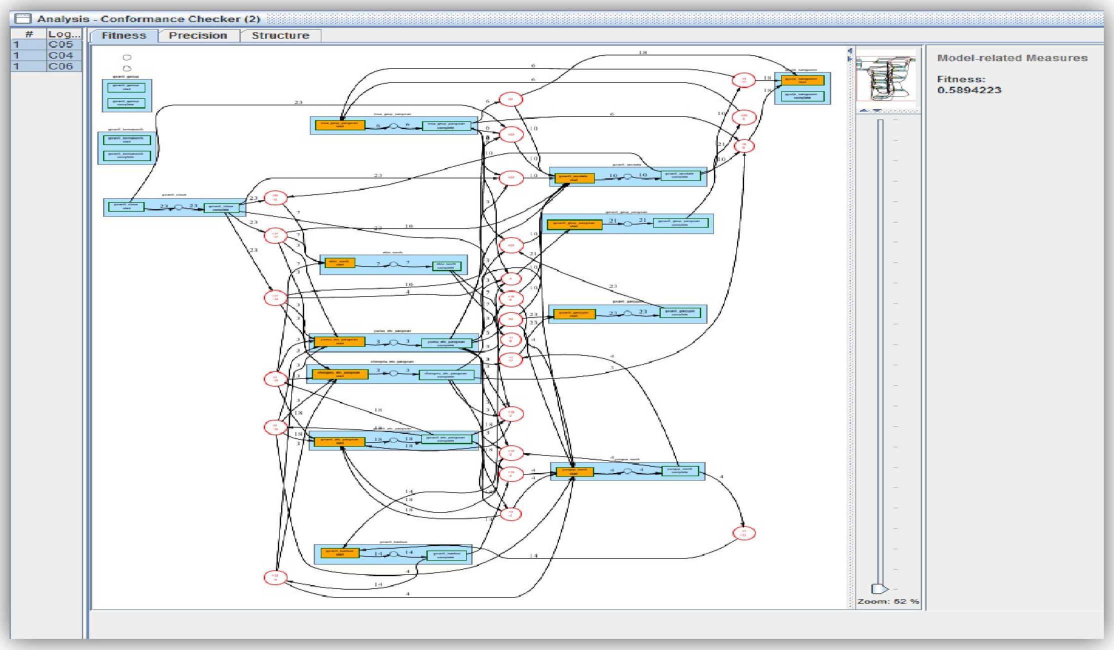


Figure 12. Conformance checking plug-in: α-mining algorithm model (Fig. 6)

Figure 13 shows the conformance checking plug-in for the heuristic mining model. For the heuristic mining model to be analyzed using the conformance checking plug-in it was converted into Petri net format prior to the analysis. The fitness metric yielded for this extracted model was found to be 81.1%. As stated earlier, the fitness also exemplifies how much more robust the heuristic mining algorithm model is in comparison to the α-mining algorithm model.

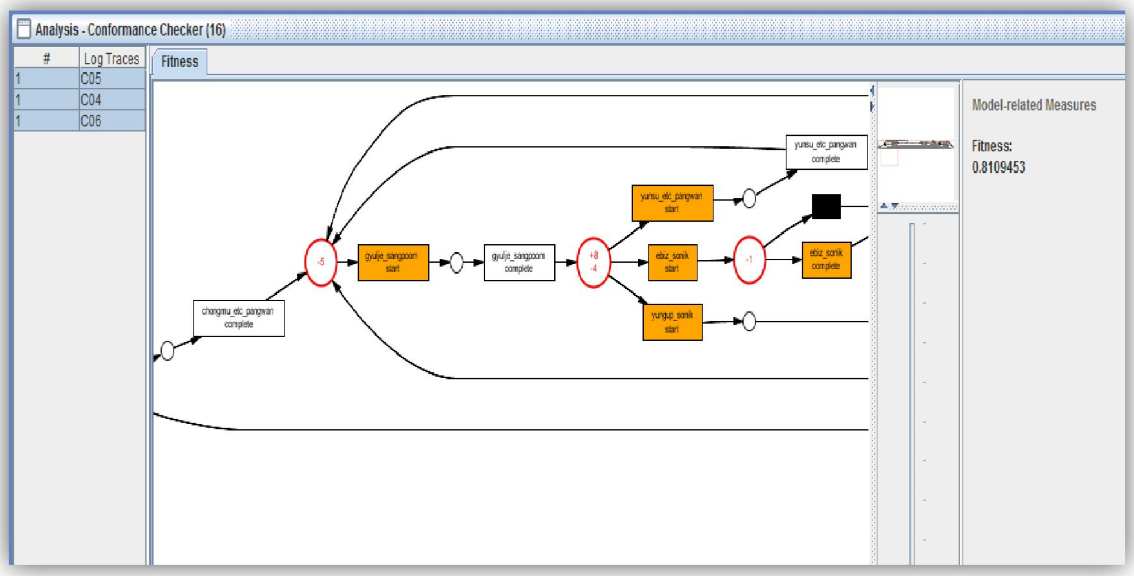


Figure 13. Conformance analysis of heuristic model (Fig. 7)

According to extracted α-mining and heuristic algorithm model (Figures 6 and 7), the start and end tasks of the process are very ambiguous, nearly undefined. As indicated by the AS-IS model, the start task should be ‘Management Accounting Preliminary Data’ (*gwanli_rawdata*) and the end task should be ‘Management Accounting Finalization’ (*gwanli_close*). Artificial start and end filters were added to locate the start and end tasks for the model extracted from the log. Additions of the artificial start and end task filters yield the start task to be defined as ‘Other Selling and General Administrative Expenses’ (*gwanli_etc_pangwan*) and the end task as ‘HQ Profit and Loss Adjustment’ (*gwanli_bonsasonik*). The creation of new models using the control flow mining techniques was repeated after the redefinition of the start and end tasks in the event log. Figures 14 and 15 represent the models with the artificial start and end tasks extracted via α-mining algorithm and heuristic mining algorithm, respectively.

It can be observed that the heuristic mining model in Figure 15 seems more robust in comparison to the α-mining algorithm model in Figure 14 due in part that there are no isolated tasks, in addition to the artificial start and end tasks being correctly applied.

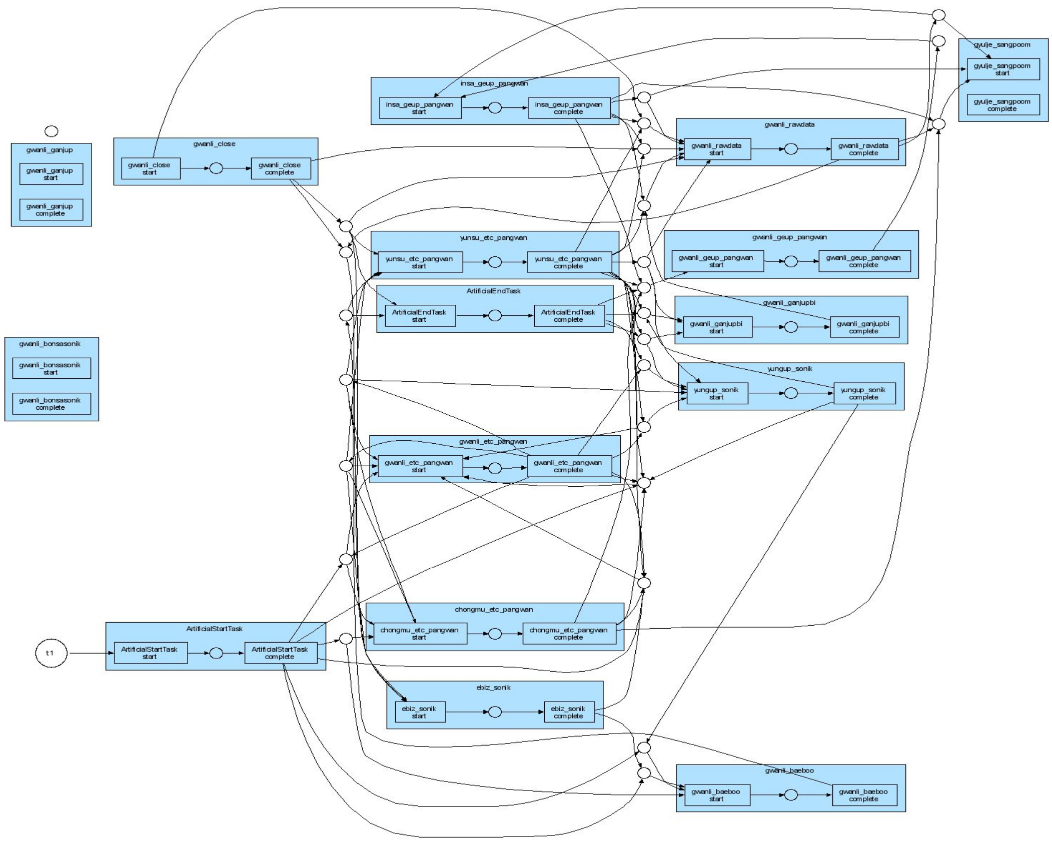


Figure 14. α-mining algorithm model with artificial start and end tasks

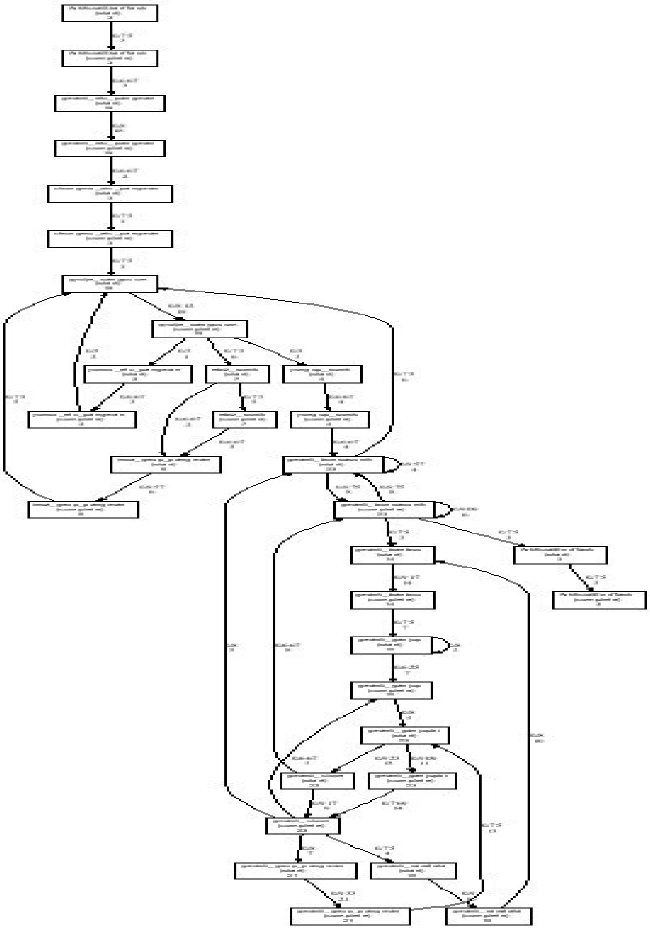


Figure 15. Heuristic mining algorithm with artificial start and end tasks

Conformance analyses of Figure 16 and 17 were performed to observe any changes. Figure 16 shows the conformance analysis of α-mining algorithm model with artificial start and end tasks.

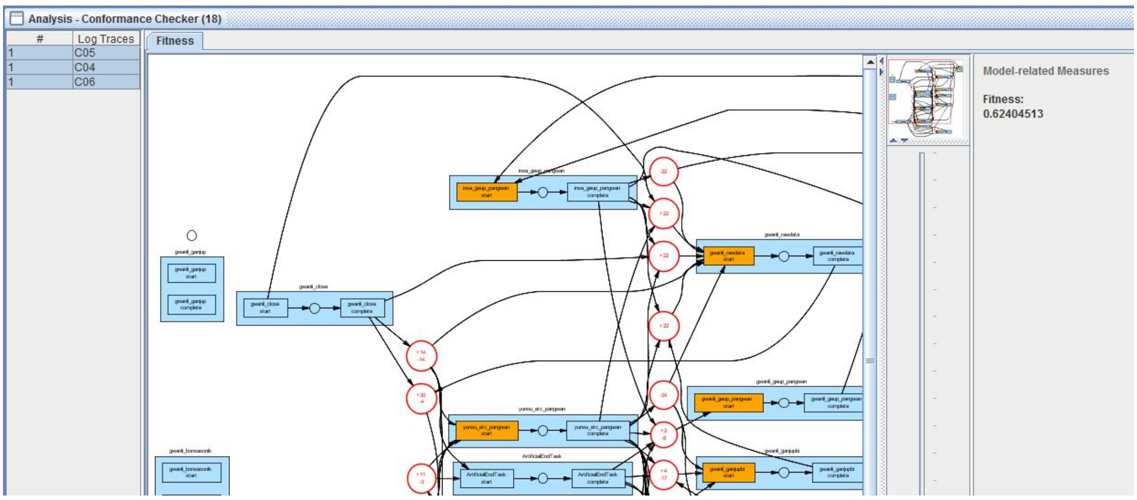


Figure 16. Conformance analysis of α-mining algorithm model with artificial start and end tasks

After adding the artificial start and end tasks to the α-mining algorithm model, the conformance was found to increase slightly from 0.589 to 0.624.

Figure 17 shows the conformance analysis of heuristic mining with artificial start and end tasks. Similar to the α-mining model with artificial start and end tasks, the fitness for the heuristic model with the artificial start and end tasks increased marginally also (0.811 to 0.830).

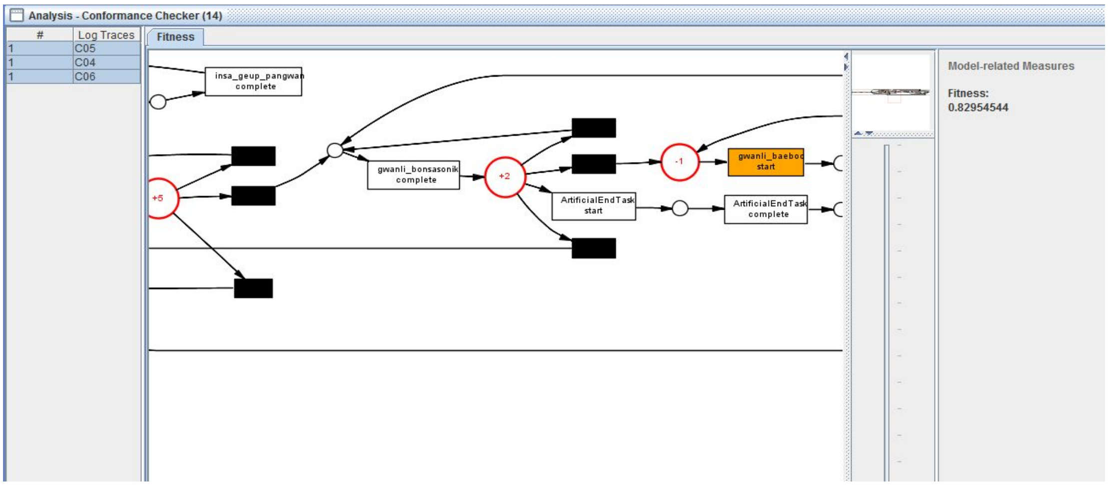


Figure 17. Conformance analysis of heuristic mining algorithm with artificial start and end tasks

Table 2 contains the summary of the conformance fitness metrics assessed in this section. It can be observed that the fitness of the AS-IS model with loops and the α-mining model are quite similar. However the fitness of the heuristic models is significantly superior to the AS-IS and α-mining models. The heuristic model is a much better representation of the actual process as recorded in the event log.

Table 2. Summary of conformance fitness metric assessment

Model	Fitness Metric (f)
AS-IS (Fig. 5)	0.402
AS-IS with Loops (Fig. 10)	0.607
α-mining (Fig. 7)	0.589
α-mining with artificial start and end tasks (Fig. 14)	0.624
Heuristic (Fig. 8)	0.811
Heuristic with artificial start and end tasks (Fig. 15)	0.830

Such discrepancies in fitness values between extracted models and the AS-IS model can be explained by the fact that the AS-IS model, prior to the addition of loops, did not account for the fact that re-executions of certain activities are present. After the addition of loops, the fitness of the AS-IS model is in par with the fitness of the α-mining algorithm model. The reasoning for this occurrence is not because the two models are structurally equivalent but they individually match the behavior observed in the log in their own respective ways.

When comparing the AS-IS model with the heuristic algorithm model in terms of fitness, we found that there exist many structural differences that may contribute to the differences in fitness values, such as the beginning and end activities, not to mention the order of activities, are significantly different between the models. Since the heuristic model is the better representation of describing the behavior of the event log as found through conformance analysis, we reveal some noteworthy observations of how processes are actually executed.

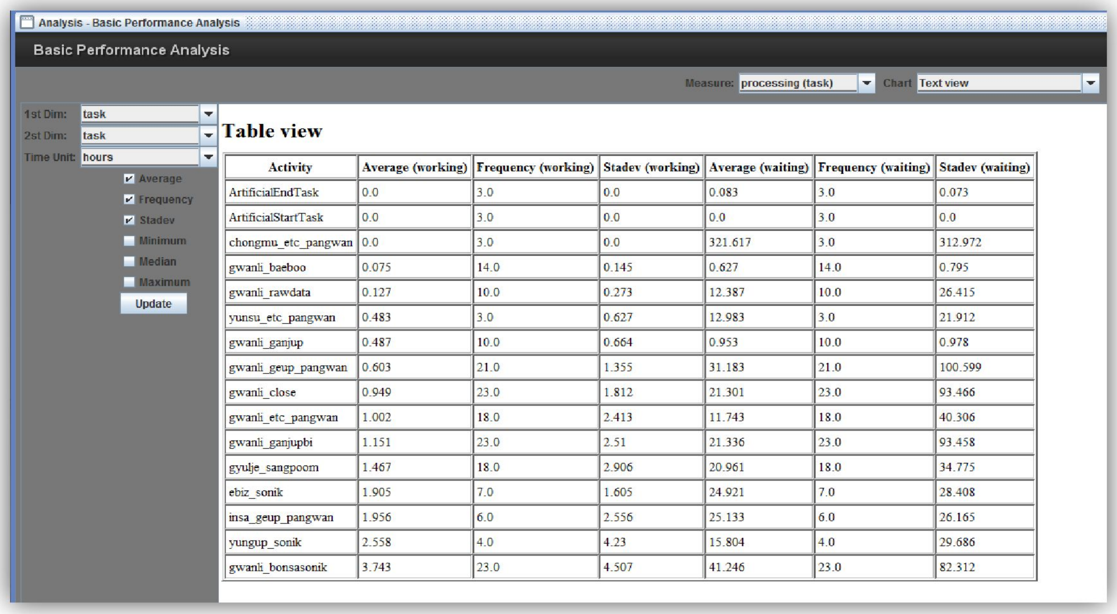
Observing the heuristic algorithm model, the start task is the *Business Management*: 'Other Selling and General Administrative Expenses' task and the end task is the *Business Management*: 'HQ Profit and Loss Adjustment' task; the AS-IS model start task, *Business Management*: 'Management Accounting Preliminary Data' is found in the middle of the process (rather than the start of the process), and is utilized more as retrieving and adjusting data for the different tasks and inter-department use; the AS-IS model end task *Business Management*: 'Management Accounting Finalization' is near the end of the process but is not the final task as it seems concluding profit and loss adjustments are made afterwards.

From a managerial accounting perspective, some of these findings were quite intriguing. First of all, from the extracted models, we found that the supportive departments (Human Resources, Payment Services, General Affairs, Training, and E-Business Departments) of the Business Management Department executed their tasks toward the earlier stages of the process. Furthermore, somewhat lesser tasks, such as the 'Other Selling and General Administrative Expenses' and 'Profit and Loss Adjustment' tasks were also performed in the earlier stages. In other words, such behaviors observed from the event logs show that the less essential tasks (or tasks that do not require much interaction with other department inputs/outputs or tasks that can be handled with ease and promptness), which generally tend to be executed by the supportive departments in regards to the Business Management Department, are completed before more vital tasks. We believe this way the Business Management and Sales Departments can utilize the preliminary data, as well as the collected data from the supportive departments, to perform the focal tasks without many obstacles and delays. However, this does not imply that the less vital tasks executed by the supportive departments are *not* revisited, as the heuristic model (Figure 15) does illustrate loops created for prior tasks to be revisited.

4.2.3 Performance Perspective

Aside from discovering process models based on event logs, process mining can contribute in many other ways such as performance analyses and assessments. This section will focus on techniques such as performance analysis to demonstrate other positive features of process mining and its supplementary value to BPR.

Basic performance analysis calculates performance measures such as execution time, waiting time, task frequencies, and many more; also it displays such measures into several types of graphs, such as bar charts and pie charts. The heuristic mining algorithm model will be analyzed since it yielded the highest fitness metric (most suitable model representing the process log) from the conformance analysis section. Figure 18 shows the results of the basic performance analysis in text view.



Activity	Average (working)	Frequency (working)	Stadev (working)	Average (waiting)	Frequency (waiting)	Stadev (waiting)
ArtificialEndTask	0.0	3.0	0.0	0.083	3.0	0.073
ArtificialStartTask	0.0	3.0	0.0	0.0	3.0	0.0
chongmu_etc_pangwan	0.0	3.0	0.0	321.617	3.0	312.972
gwanli_baeboo	0.075	14.0	0.145	0.627	14.0	0.795
gwanli_rawdata	0.127	10.0	0.273	12.387	10.0	26.415
yunsu_etc_pangwan	0.483	3.0	0.627	12.983	3.0	21.912
gwanli_ganjup	0.487	10.0	0.664	0.953	10.0	0.978
gwanli_geup_pangwan	0.603	21.0	1.355	31.183	21.0	100.599
gwanli_close	0.949	23.0	1.812	21.301	23.0	93.466
gwanli_etc_pangwan	1.002	18.0	2.413	11.743	18.0	40.306
gwanli_ganjupbi	1.151	23.0	2.51	21.336	23.0	93.458
gyulje_sangpoom	1.467	18.0	2.906	20.961	18.0	34.775
ebiz_sonik	1.905	7.0	1.605	24.921	7.0	28.408
insa_geup_pangwan	1.956	6.0	2.556	25.133	6.0	26.165
yungup_sonik	2.558	4.0	4.23	15.804	4.0	29.686
gwanli_bonsasonik	3.743	23.0	4.507	41.246	23.0	82.312

Figure 18. Basic Performance Analysis

Often performance analysis is performed to obtain two major performance measures: a) measures for tasks (processing), and b) measures for instances (throughput). We conducted basic performance analysis prior to performance analysis to obtain a general schema of the performance measures of the process. We found that the ‘HQ Profit and Loss Adjustment’ task performed by the Business Management department had the longest working hours, and the ‘Other Selling and General Administrative Expenses’ task conducted by the General Affairs department had the longest average waiting time.

Moreover performance analysis was performed to acquire a more in-depth view into the process. Performance analysis uses the log replay method, which simulates the process instances in the input log in the model. In order for the log replay method to work, the events in the log have to be associated with the traces in the model. For this reason we utilize the model with the highest fitness metric, the heuristic algorithm model. Figure 19 shows the results from the performance analysis. The places are distinguished into three different colors that signify the low (blue), medium (yellow), high (purple) levels of waiting time, thus this is a visual depiction of bottlenecks in the process simulated on the model.

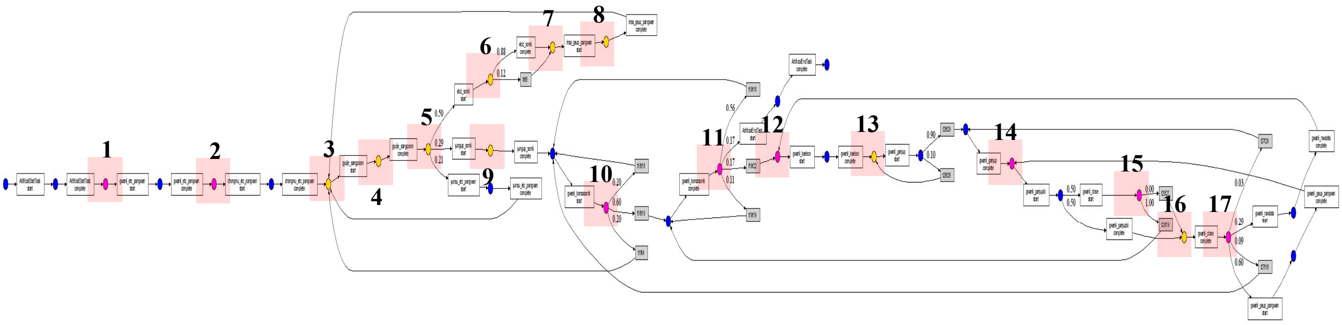


Figure 19. Performance analysis of heuristic mining algorithm model

Table 3 was created to organize the medium to high level bottlenecks found through performance analysis. Overall there exist nine medium level bottlenecks and eight high level bottlenecks. The average execution time denotes the duration of a task from start to completion; the average waiting time signifies the waiting time before a certain task is performed; and the average sojourn time indicates the total time of the waiting time and synchronization time. The synchronization time is defined as the time that passes from the partial enabling of a transition (i.e., transition with more than one input place) until it is fully enabled, thus all input places of the transition must be marked for the transition to “fire”.

Table 3. Summary measures of average execution, waiting, and sojourn times of medium to high level bottlenecks

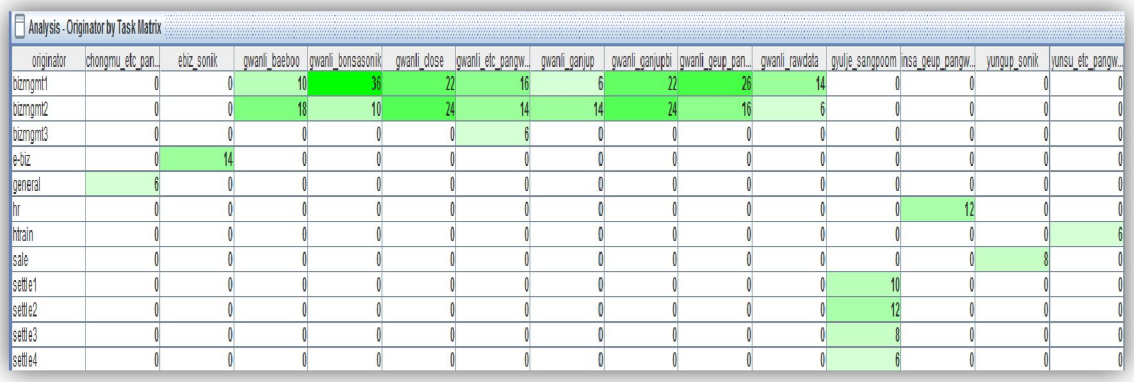
Places: Bottleneck *(Red Highlighted Squares; Fig.20)*	Average Execution Time (hrs)	Average Waiting Time (hrs)	Average Sojourn Time (hrs)
Within Activity			
4	1.47	-	-
6	1.41	-	-
8	1.68	-	-
9	3.41	-	-
Between Adjacent Activities			
1	-	96.34	96.34
2	-	469.83	469.83
3	-	18.63	18.63
5	-	35.53	35.53
7	-	33.79	33.79
10	-	62.18	62.18
11	-	144.01	144.01
12	-	35.65	35.65
13	-	5.44	5.44
14	-	42.8	42.8
15	-	143.25	143.25
16	-	1.64	1.64
17	-	167.39	167.39

According to Table 3, we can identify the high level bottlenecks and their average waiting times. The eight high level bottlenecks occur *before* the following tasks: a) *Business Management*: 'Other Selling and General Administrative Expenses', b) *General Affairs*: 'Other Selling and General Administrative Expenses', c) *Business Management*: 'HQ Profit and Loss Adjustment', d) *Business Management*: 'Capital Cost Distribution', e) *Business Management*: 'Overhead Cost Allocation/Finalization', f) *Business Management*: 'Management Accounting Finalization', g) *Business Management*: 'Management Accounting Preliminary Data', and h) *Human Resources*: 'Wage Expense and Other Selling and General Administrative Expenses Management'. Such high level bottlenecks, as well as the medium level bottlenecks, are identified and recommended for further analyses to the company. Through process mining, the diagnosis and detection of such problems can help the company undergo more efficient, effective and significant redesign of the process. As a side note, bottlenecks 1 and 11 can be omitted due to the additions of the artificial start and end tasks into the event log.

4.2.4 Organizational and Social Network Perspective

Our final analyses pertain to the organization structure and social network aspects of the company, using the event log. Such analysis is more commonly known as organizational mining (Song and van der Aalst, 2008), where the organizational setting and interactions among coworkers are extracted from event logs. Organizational models generated from organizational mining can assist in improving the underlying processes.

As a foundation, the originator by task matrix plug-in was utilized to show the relationship between tasks and originators, shown in Figure 20. Also from the matrix, information regarding the existence of sub-groups in the departments and the tasks they perform can be obtained. From Figure 20, the Business Management department is categorized into three different sub-groups and the Payment Services department has four sub-groups. Furthermore the task matrix color coordinates the frequencies of the cells with different color intensities. The frequencies of the tasks in regards to sub-groups can help recognize specialized roles. For example the third group of the Business Management Department specializes in tasks pertaining to 'Other Selling and General Administrative Expenses'.


Analysis - Originator by Task Matrix

originator	chongmu_etc_pan...	ebiz_sonik	gwanli_baeboo	gwanli_bonsasonik	gwanli_close	gwanli_etc_pangw...	gwanli_ganjup	gwanli_ganjupbi	gwanli_geup_pan...	gwanli_rawdata	gyulje_sangpoom	insa_geup_pangw...	yungup_sonik	yunsu_etc_pangw...
bizmgmt1	0	0	10	36	22	16	6	22	26	14	0	0	0	0
bizmgmt2	0	0	18	10	24	14	14	24	16	6	0	0	0	0
bizmgmt3	0	0	0	0	0	6	0	0	0	0	0	0	0	0
e-biz	0	14	0	0	0	0	0	0	0	0	0	0	0	0
general	6	0	0	0	0	0	0	0	0	0	0	0	0	0
hr	0	0	0	0	0	0	0	0	0	0	0	12	0	0
htrain	0	0	0	0	0	0	0	0	0	0	0	0	0	6
sale	0	0	0	0	0	0	0	0	0	0	0	0	8	0
settle1	0	0	0	0	0	0	0	0	0	0	10	0	0	0
settle2	0	0	0	0	0	0	0	0	0	0	12	0	0	0
settle3	0	0	0	0	0	0	0	0	0	0	8	0	0	0
settle4	0	0	0	0	0	0	0	0	0	0	6	0	0	0

Figure 20. Originator by task matrix

Moreover, the handover of work using social network analysis was performed, shown in Figure 21. The basic idea behind the handover of work is that performers are related if there is a causal relation through the passing of a case from one performer to another (van der Aalst et. al., 2004). From Figure 21, we can observe that 'bizmgmt1' and 'bizmgmt2' play a central role in general, where they possess causal interactions with the majority of the supportive departments. Also the Payment Services Department (settle1, settle2, settle3, settle4) is divided into four sub-groups, which partake in the handover of work with all departments with the exception of 'bizmgmt3', 'hr', and 'sale' departments.

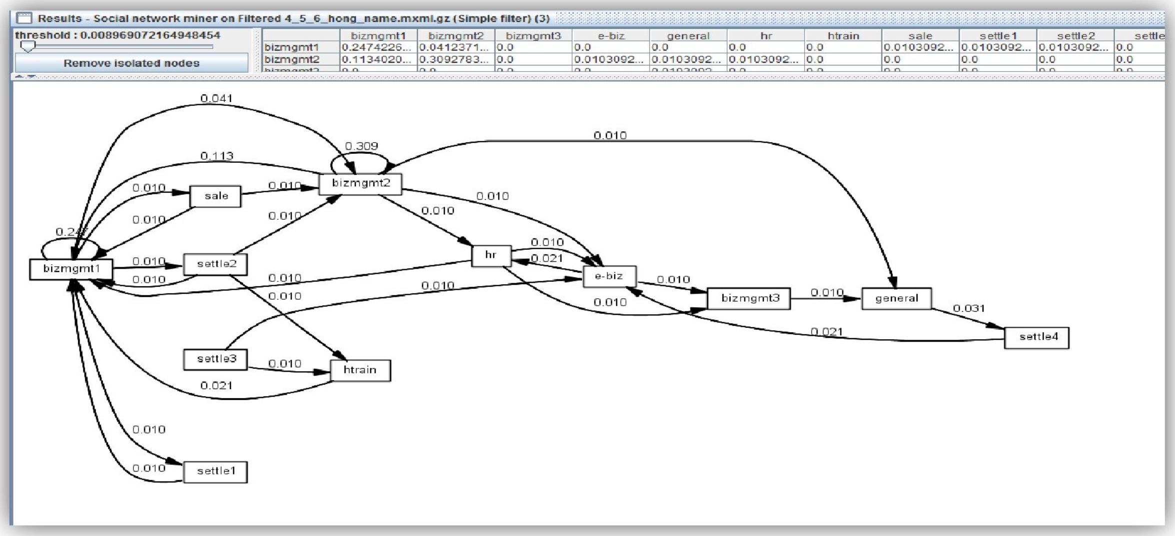


Figure 21. Social Network Miner

V. Discussion

Our research was motivated by adding to the current process mining application literature by examining the applicability of process mining techniques in a new area, the financial sector. From the previous section we performed process mining techniques from three different perspectives: the control flow perspective, performance perspective, and organizational and social network perspective. Also a conformance analysis was conducted to further investigate the process models.

From the control flow perspective, our results discovered that the AS-IS model was not an accurate representation of the actual behavior recorded in the event log, nor was the process as simple and straight-forward as it was illustrated in the AS-IS model. We learned that AS-IS models tend to describe the ideal flow of the process under the assumption that all tasks are performed correctly without any flaws or mishaps and the general sequence of activities in the process. In other words, by comparing the AS-IS model with the extracted models from the event log, we found that the actual process contained repeated tasks, or loops, not to mention revisits of previously conducted or completed tasks earlier in the process; thus the actual behaviors of the process were quite complex.

To further investigate from a control flow perspective, conformance analysis was performed to obtain better insights of the AS-IS model and the extracted models. Since the AS-IS model can be accepted as a representation of the ideal process assuming nothing goes wrong, it is understandable that it lacks in conformance to the actual log of the process. Of the discovered models extracted from process mining, the α-mining algorithm model yielded low fitness measures with similar values to the AS-IS model. This similarity does not imply that the α-mining model is unreliable, but that, since α-mining algorithm maps out every activity in the event log, the model is subject to weaknesses such as noises, as it is a direct and verbatim visual representation of the event log as it was recorded. In other words, it extracts too much information to gain intriguing insights of the actual process. Furthermore the similarities of the fitness values of the AS-IS model and the α-mining model does not imply that the two models are structurally equivalent but they individually match the behavior observed in the log in their own respective ways.

On the other hand, heuristic mining algorithm compensates for some of α-mining algorithm's weaknesses (i.e., noises in the log) in that heuristic algorithm accounts for frequencies and relational strengths between the tasks in order to focus on the significant traces of the process log. As a result, we found that the heuristic mining model yielded the highest fitness value, thus it being the closest model to match the behavior of the process according to the event log. By comparison, the AS-IS model and the heuristic mining model had distinct and significant differences. From the differences in the tasks that were the starting (and ending) tasks of the process to the differences in sequential orders of the process, it was hard to find any evidence that the two models were "describing" the same process. Moreover from the heuristic mining model, we observed that the supportive departments

(Human Resources, Payment Services, General Affairs, Training, and E-Business Departments), whom are responsible for the less vital tasks (wage expense, sectional and product profits and loss computations as well as other selling and general administrative expense computations) of the monthly profit and loss computation process, were performed earlier in the process, then followed by the more important tasks which were conducted by the Business Management Department. Not only is this observance evidence of the allocations of work being dispersed among the supportive departments, thereafter their outputs being collected for the important core tasks to be performed by the Business Management Department, but also this provides a closer insight as to how the process transpires. The AS-IS model illustrates that all supportive department tasks are performed in parallel to each other as well as the tasks performed in the Business Management Department, however we can see that there is some structure in the order of which the tasks are being performed, and further that such knowledge can be of some assistance to the management when redesigning this process.

From the performance perspective of our research, we found the existence of bottlenecks within the process. One key advantage of utilizing process mining as a supplementary tool in BPR is that problems pertaining to performance within the process can be pinpointed visually with ease, which was previously difficult to locate with its traditional methods of process assessment. Through performance analysis, the specificities of the bottleneck can be identified according to which task is being problematic, who is performing the task, and how severe the problem is in of itself and to the process as a whole. For example in reference to Figure 19 and Table 3, a high level bottleneck, bottleneck 2, has the highest and most extensive waiting time of 469.83 hours among all other bottlenecks. Bottleneck 2 represents the phase between the *Business Management*: 'Other Selling and General Administrative Expenses' task and the *General Affairs*: 'Other Selling and General Administrative Expenses' task. We suggest that further inspection of activities or problems between the two departments is highly recommended. The issues that may be causing the high level bottleneck may stem from problems in communication between departments, employee capabilities, middle to lower management, etc. A complete and viable solution in fixing such bottlenecks is difficult to generate with current process mining technique alone.

In terms of the organizational and social network perspective of our research, in our opinion, there were no staggering details that were worthy of scrutiny or special attention. Perhaps due to the lack of knowledge of how inter-departmental relationships and sub-groups within departments interact, there exists some difficulty in obtaining suggestion in which to help how certain organizational groups should be cooperating amongst each other. Furthermore due to most tasks being performed by the Business Management Department, it is quite difficult to locate qualms between different departments, since the supportive departments execute only one task per department. To obtain further insight of the organizational social network, performing analyses at the individual or the group/team level of analysis is suggested.

Overall, the heuristic algorithm model demonstrates that process mining can provide insight into how processes are actually executed. Furthermore, because process mining utilizes concrete data from information systems logs, it is essentially immune to the problem of misleading or missing information when assessing or diagnosing the status of an organization through interviews and surveys.

Another goal of our research was to accentuate some of the strengths of process mining and how process mining can be a *supplementary* tool in enhancing the current methodologies of BPR, thus in hopes to eventually and indirectly increase its success rates. Table 4 organizes the strengths of process mining in a practical setting regarding data, process, performance, and organizational assessment methods, in which can aid in areas in which BPR lacks.

Table 4. Comparison of Process Mining and BPR

	Process Mining	BPR
Data Assessment Methods	Concrete process log data • Identifiable activity, originator, case ID, time stamp information	Interviews and Surveys • Subject to biases, misleading, and lack of information • Time consuming
Processes Discovery	Automated discovery of process model based on log behavior • Easy extraction of process model • Numerous algorithms to find the best model that represents log behavior • Ability to test the fitness of the extracted models	Interviews and Surveys • Time consuming • Subject to loss of information due to turnover • Difficulty in recall of past activities and information
Performance Assessment	Detailed performance measures and locating bottlenecks • Pinpoint precision in locating problems in the process • Ability to assess waiting and execution times in specific locations in the process	Performance Assessment Tools • Interview/Survey biases of employee self-evaluations of performance • Performance Work Statement • Workforce Statistics • Equipment and Material Assessment
Organizational Perspective	Automated multiple visualizations of organizational structure and social networks • Subcontracting & reassignment • Handover of work between individuals and departments • Similar tasks of originators	Interviews and Surveys • Difficulty assessing the handover of work between individuals and departments

VI. Conclusion

In conclusion, the purpose of this study was to accomplish the following: a) add to the current process mining application literature by examining the applicability of process mining in the financial sector, b) utilize various process mining techniques to observe and assess process information in the specific managerial accounting case study of the monthly profit and loss computation process, c) identify the strengths of process mining and how they can supplement the weaknesses of business process reengineering, and d) address the possibility that event log data with insufficient case sizes needs to be addressed differently than past approaches with data with sufficient case sizes.

In this study, we have focused on the applicability of process mining in the financial domain. As a rapidly emerging field, the importance of the applicability of process mining over vast fields and domains cannot be disregarded. First this study was motivated by the lack of an application of process mining in the financial sector. We have shown the possibility of applying process mining to a complex process, obtaining insights into the process, and deriving reasonable models for a process with limited cases.

For the case study, we have used data obtained from a monthly profit and loss computation process of an investments and securities firm. Unlike most prior research of the applicability of process mining of structured processes, the monthly profit and loss computation process of the firm was found to be complex. This process was also found to be intricate in a sense that the process was not as straight-forward as company experts had envisioned, and had many activities that were being performed at the same time by different departments within the company. Also the actual process contained many repetitions of tasks and tasks that were revisited after their completion.

Furthermore we have identified the strengths of process mining and such attributes can aid in the context of business process reengineering. For example, process mining allows for the discovery of actual processes into models and beneficial visualizations, as well as process diagnosis, where the existing process is analyzed to identify problems, or to find areas for improvement. Our research does not imply that process mining will substitute BPR, but emphasize that it can provide significant benefits in BPR implementation.

Limitation and Future Directions

There exist some limitations in our research. A clear distinction between the existing model and the model extracted from the control flow mining techniques meant one of two things – whether currently existing process mining techniques are appropriate in the assessment and analyses of extracting proper models from the event log or the AS-IS model is inappropriately conceptualizing how the process actually transpires. The main limitation of this study is the lack of methods in which to evaluate the validity of the process mining results. This is due in part of the insufficiency of existing evaluation techniques in the process mining domain (Rozinat et. al., 2008). Rozinat et. al. (2008) have already claimed the urgency of a process mining evaluation framework that enables a) process mining researchers to compare the performance of their algorithms, and b) end users to evaluate the validity of their process mining results. This lack of evaluation techniques triggers the motivation to construct such evaluation methods or a framework for process mining results in future research, as well as to further strengthen current process mining tools and originate more robust algorithms for future endeavors.

Another specific limitation to this study was the low case size within the data. Since the event log data consisted of the records of the monthly profit and loss computation process, each case represented the events of the process within a single month. Similar to the field of statistics, the more data (or more specifically, cases) that are available, the more accurate the analyses of the data can represent and examine the actual process and the analyses can be less vulnerable to noises and deviations within the data. Unfortunately the data obtained only contained three cases, or three months. Although the assurance that the cases reflecting the actual process can be increased by obtaining more cases, unlike the statistics domain where much research in terms of calculating acceptable sample and data size have been accomplished, the process mining domain is still growing and, in our opinion, is in dire need of research of specifying how many cases is required for proper process mining analysis to be conducted. Furthermore in terms of case sizes, exceptions need to be made for applications of process mining research for low case sizes due to the fact not all event log data have the abundance of cases. Different from statistics, increasing sample (case) size tends to be more difficult, if not impossible for event log data, thus we also present the need of a framework for process mining research and applications for small case sizes.

References

1. Adriansyah, A., van Dongen, B.F., and van der Aalst, W.M.P. 2010. 'Towards Robust Conformance Checking'. BPM'10 Workshops, vol. 66, pp. 122-133.
2. Alves de Medeiros, A.K. 2006. Genetic Process Mining. *PhD thesis*, Eindhoven University of Technology, Eindhoven.
3. Arkin, A., Askary, S., Bloch, B., Curbera, F., Goland, Y., Kartha, N., Liu, C.K., Thatte, S., Yendluri, P., and Yiu, A. 2005. 'Web Services Business Process Execution Language Version 2.0.' WS-BPEL Technical Report, OASIS.
4. Athena, Pallas. 2002. 'Flower User Manual'. Pallas Athena BV, Apeldoorn, The Netherlands.
5. Boumen, R., de Jong, I.S.M, Vermunt, J.W.H., van de Mortel-Fronczak, J.M., and Rooda, J.E. 2006. 'Test Sequencing in Complex Manufacturing Sytems', *IEEE Transactions on System, Man, and Cybernetics* – Part A: 39(1), pp. 177-187.
6. Cook, J.E., Wolf, A.L. 1999. 'Software Process Validation: Quantitatively Measuring the Correspondence of a Process to a Model'. *ACM Transactions on Software Engineering and Methodology* 8(2), pp. 147–176.
7. Dumas, M., van der Aalst, W.M.P., and ter Hofstede, A.H.M. 2005. 'Process-Aware Information Systems: Bridging People and Software through Process Technology'. Wiley & Sons, Chichester.
8. Funk, M., Rozinat, A., Alves de Medeiros, A.K., van der Putten, P.H.A., Corporaal, H., and van der Aalst W.M.P. 2009. *J. Yang (Ed.), UNISCON 2009. (Lecture Notes in Business Information Processing, Vol. 20, pp. 190-201). Berlin: Springer.*
9. Garrison R. & Noreen E. 2009. *What is Managerial Accounting. Accounting for Mangement Website*.[Online] Available at: www.accountingformanagement.com/what_is_managerial_accounting.htm [Accessed May 15, 2011].
10. Hammer, M. and Champy, J. 1993. *Reengineering the Corporation*. Nicolas Brealey Publishing, London.
11. Kavantzas, N., Burdett, D., Ritzinger, G., Fletcher, T., and Lafon, Y. 2004. *Web Services Choreography Description Language Version 1.0* [Online] Available at: http://www.w3.org/TR/2004/WD-ws-cdl-10-20041217/ [Accessed March 27, 2011].
12. Larsen, M. and Myers, M. 1997. 'BPR Success or Failure?: A Business Process Reengineering Project in the Financial Services Industry', ICIS '97 Proceeding. Paper 24.
13. Mans, R.S., Schonenberg, M.H., Song, M., van der Aalst, W.M.P, and Bakker, P.J.M. 2008. 'Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital', *BIOSTEC 2008*, CCIS 25, pp. 425-438.

14. Maruster, L., van der Aalst, W.M.P., Weijters, A.J.M.M., van den Bosch, A., and Daelemans, W. 2002. 'Automated Discovery of Workflow Models from Hospital Data', *Proceedings of the ECAI Workshop on Knowledge Discovery and Spatial Data,* pp. 32-36.
15. Maruster, L., Wortmann, J.C., Weijters, A.J.M.M., and van der Aalst, W.M.P. 2002. 'Discovering Distributed Processes in Supply Chains', *Proceedings of the International Conference on Advanced Production Management Systems (APMS 2002),* pp. 119-128.
16. Rozinat, A., de Jong, I.S.M., Günther, C.W., and van der Aalst, W.M.P. 2009. 'Process Mining Applied to the Test Process of Wafer Steppers in ASML', IEEE Transactions on Systems, Man, and Cybernetics, Part C, 39(4), pp. 474-479.
17. Rozinat, A., Alves de Medeiros, A.K., Günther, C.W., Weijters, A.J.M.M., and van der Aalst, W.M.P. 2008. 'The Need for a Process Mining Evaluation Framework in Research and Practice', *BPM 2007 Workshop, Lecture Notes in Computer Science* vol. 4928, pp. 84-89.
18. Rozinat, A., de Jong, I.S.M., Günther, C.W., and van der Aalst, W.M.P. 2009. 'Conformance Analysis of ASML's Test Process'. *Proceedings of the Second International Workshop on Governance, Risk and Compliance (GRCIS'09)* vol. 459, pp. 1-15.
19. Song, M. and van der Aalst, W.M.P. 2008. 'Towards Comprehensive Support for Organizational Mining'. *Decision Support Systems* 46(1), pp. 300-317.
20. Song, M., Günther, C.W. and van der Aalst, W.M.P. 2008. 'Trace Clustering in Process Mining', *4th Workshop on Business Process Intelligence,* BPI 08*,* pp. 51-62.
21. Song, M., Kang, Y., Ryu, C., 2010. '프로세스 마이닝을 활용한 프로세스 분석 방법론의 개발과 활용: 공공부문 중심으로'. 2010 년 한국산업정보학회 춘계학술대회, UNIST, 2010.
22. van der Aalst, W.M.P., Reijers, H.A., van Dongen, B.F., Herbst, J., Maruster, L., Schimm, G., Weijters, A.J.M.M. 2003. 'Workflow Mining: A Survey of Issues and Approaches'. *Data and Knowledge Engineering* 47(2), 237-267.
23. van der Aalst, W.M.P. and Song, M. 2004. 'Mining Social Networks: Uncovering Interaction Patterns in Business Processes', *Computer Science*, vol. 3080, pp. 244-260.
24. van der Aalst, W.M.P., Weijters, A.J.M.M., Maruster, L. 2004. 'Workflow Mining: Discovering Process Models from Event Logs'. *IEEE Transactions on Knowledge and Data Engineering* 16(9), pp. 1128-1142.
25. van der Aalst, W.M.P., Reijers, H.A., Weijters, A.J.M.M., van Dongen, B.F., Alves de Medeiros, A.K., Song, M., & Verbeek, H.M.W. 2007. 'Business Process Mining: An Industrial Application'. Information Systems 32(5), pp. 713-732.
26. van der Aalst, W.M.P. and Günther, C.W. 2007. 'Finding Structure in Unstructured Processes: The Case for Process Mining', *Applications of Concurrency to System Design*, pp. 3-12.

27. van der Aalst, W.M.P., van Hee, K., van der Werf, J.M., and Verdonk, M. 2010. ‘Auditing 2.0 Using Process Mining to Support Tomorrow’s Auditor’. *IEEE Computer* 43(3), pp. 90–93.
28. van der Aalst, W.M.P., van Hee, K., van der Werf, J.M., Kumar, A., and Verdonk, M. 2011. ‘Conceptual Model for Online Auditing’. *Decision Support Systems* 50, pp. 636-647.
29. van Dongen, B.F. & Adriansyah, A. 2009. ‘Process Mining: Fuzzy Clustering and Performance Visualization’. *5th Workshop of Business Process Intelligence*, BPI 09. vol. 43, pp. 158-169.
30. Weijters, A.J.M.M. and van der Aalst, W.M.P. 2003. ‘Rediscovering Workflow Models from Event-Based Data using Little Thumb’. *Integrated Computer-Aided Engineering* 10(2), pp. 151–162.
31. Wen, L.,Wang, J., Sun, J.G. 2006. ‘Detecting Implicit Dependencies Between Tasks from Event Logs’. *APWeb 2006*. Lecture Notes in Computer Science, vol. 3841, pp. 591–603.
32. Wen, L.,Wang, J., van der Aalst, W.M.P., Huang, B., and Sun, J.G. 2010. ‘Mining Process Models with Prime Invisible Tasks’. *Data & Knowledge Engineering* 69(10), pp. 999-1021.

Acknowledgement

First of all, it is with immense gratitude that I acknowledge the support, benevolence, and most of all, patience of my thesis and graduate advisor at UNIST, Dr. Minseok Song. Without his undivided dedication, his undeviating guidance, and continual encouragement, this thesis would have been a farfetched dream. My thanks and appreciation goes to my thesis committee members, Dr. Kwanho Kim and Dr. Han-Gyun Woo, for their help and support regarding this thesis. Also I would like to thank each and every professor within the School of Technology Management, Dr. Jin-Hyouk Im, Dr. Kwanho Kim, Dr. Kyootai Lee, Dr. Insook Cho, Dr. Minseok Song, Dr. Han-Gyun Woo, Dr. Dongryul Lee, Dr. Yoonhyuk Jung, Dr. Wooje Cho, and Dr. Daniel Shim; it was a true pleasure and great honor to work and learn from the some of the best of each discipline. Moreover I would like to acknowledge Dr. Jae-Hong Park with thankfulness for his generosity and hospitality during his time present at UNIST.

I would like to thank and acknowledge my department colleagues for helping me endure through both academic and personal struggles throughout the two years of our M.S. program. I would like to gratefully recognize the ‘members’ of UMGs, Jeong-il, Jong Sang, Young Sim, Mihee, Seo-Yoon, Myoung Hee, and Heejin. I am truly blessed to find such unique and loving individuals who have helped me improve and strive to become a better person, not to mention their loyal and supportive presence to help me endure through my lowest times. Also they were often my sole source of joy and happiness, creating my graduate environment into an atmosphere where many others would only envy. I cannot find the words to even partially express my special gratitude toward Jeong-il, who stood beside me and encouraged me since the day we met. I sincerely thank you for being so compassionate and patient with me, for I am forever in debt for all you have sacrificed and done for me.

Last, but definitely not least, I dedicate this thesis to my family, for their guidance, mental support, and their belief in me throughout my journey of life thus far. I would not be standing where I stand today without their devoted love and passion. Every time I was deterred physically, mentally, and emotionally, it was from their affection and encouragement where I found the strength to persevere.